Leveraging Parameter Space Symmetries for Reasoning Skill Transfer in LLMs

Stefan Horoi[1,2,†,*], **Sangwoo Cho**[3], **Supriyo Chakraborty**[3], **Shi-Xiong Zhang**[3], **Sambit Sahu**[3], **Guy Wolf**[1,2], **Genta Indra Winata**[3,†]
[1]Université de Montréal [2]Mila – Quebec AI Institute [3]Capital One
stefan.horoi@mila.quebec, genta.winata@capitalone.com

Editors: Marco Fumero, Clementine Domine, Zorah Lähner, Irene Cannistraci, Bo Zhao, Alex Williams

Abstract

Task arithmetic is a powerful technique for transferring skills between Large Language Models (LLMs), but it often suffers from negative interference when models have diverged during training. We address this limitation by first aligning the models' parameter spaces, leveraging the inherent permutation, rotation, and scaling symmetries of Transformer architectures. We adapt parameter space alignment for modern Grouped-Query Attention (GQA) and SwiGLU layers, exploring both weight-based and activation-based approaches. Using this alignment-first strategy, we successfully transfer advanced reasoning skills to a non-reasoning model. Experiments on challenging reasoning benchmarks show that our method consistently outperforms standard task arithmetic. This work provides an effective approach for merging and transferring specialized skills across evolving LLM families, reducing redundant fine-tuning and enhancing model adaptability.

1 Introduction

The proliferation of open-weight Large Language Models (LLMs) [9, 15, 3, 21] has fostered a decentralized ecosystem of model development. Once a foundational model is released, research teams independently adapt it to diverse applications. However, the rapid release of next-generation base models introduces a developmental dilemma: teams must either continue relying on their older, customized models or incur the substantial cost of re-adapting their work to a newer, more capable version. This challenge underscores the need for methods that can efficiently integrate the specialized skills of an adapted model with the enhanced capabilities of its successor.

A promising strategy for skill transfer is task arithmetic [14], which represents the effect of fine-tuning as a task vector, the difference between the fine-tuned and pre-trained weights. This vector encodes the acquired skill, and by adding multiple task vectors to a base model, their respective skills can be combined [14]. Despite its appeal, task arithmetic is prone to negative interference, where conflicting parameter updates from independently trained models degrade performance [23, 24]. This problem becomes especially severe when models have substantially diverged during adaptation.

This paper addresses two challenges at this intersection: 1) efficiently transferring skills across different versions of a foundational model, and 2) mitigating the negative interference that arises when applying task arithmetic to models with diverging training trajectories. We focus on the crucial skill of LLM reasoning. Cultivating robust reasoning in LLMs is a notoriously difficult and

[*]The work was done while in an internship at Capital One. [†]Corresponding authors.

Proceedings of the III edition of the Workshop on Unifying Representations in Neural Models (UniReps 2025).

computationally expensive process. Therefore, the ability to transfer these hard-won reasoning skills between model versions is a vital research objective, allowing teams to leverage state-of-the-art advancements without invalidating their prior investments.

Contributions. We make several contributions. First, we extend methods that leverage parameter space symmetries, specifically rotation and scale symmetries for attention layers [25] and permutation symmetry for feed forward layers (FFN) [2], to their widely used, modern counterparts, Grouped-Query Attention (GQA) [1] and SwiGLU layers [18]. Second, we use these alignment methods to transfer the reasoning skills of `Nemotron-Nano` [3] from its reference model, `Llama-3.1-8B-Instruct` [9], to the independently instruction-tuned `Tulu3-8B` model. Through extensive evaluations, we show that aligning the parameter spaces before transferring the reasoning skills using task arithmetic significantly improves the final model's performance on hard reasoning benchmarks.

2 Preliminaries and Methodology

2.1 Task Arithmetic for Skill Transfer

Task arithmetic [14] provides an efficient framework for editing model capabilities by performing linear operations directly on their weights (θ). The core idea is to represent a learned skill as a **task vector** (τ_{skill}), which is the difference between a fine-tuned model's parameters and its original base model's parameters. This vector isolates the changes that encode a specific new skill.

$$\tau_{\text{skill}} = \theta_{\text{fine-tuned}} - \theta_{\text{base}}. \tag{1}$$

This formulation is particularly powerful for **skill transfer**. A skill vector, such as one for reasoning, can be extracted from a reference model pair (e.g., $\tau_{\text{reason}} = \theta_{\text{ref.+reason}} - \theta_{\text{ref.}}$) and then applied to an entirely different target model (θ_{target}) to imbue it with that same skill. The new model's parameters, $\theta_{\text{target+reason}}$, are synthesized through simple vector addition:

$$\theta_{\text{target+reason}} = \theta_{\text{target}} + \tau_{\text{reason}}. \tag{2}$$

This creates an analogy where the target model is enhanced in the same way as the reference model. While effective, this process can suffer from **negative interference** [23, 24], where conflicting parameter updates between the models degrade performance, especially when their parameter spaces have significantly diverged.

2.2 Leveraging Parameter Space Symmetries in Neural Networks

The architecture of modern neural network are characterized by multiple parameter space symmetries, meaning distinct sets of weights can produce an identical output function. For instance, neurons in a feed-forward layer can be reordered, or the internal representations in an attention layer can be rotated, without changing the model's behavior. These misalignment can occur when models are trained from different initializations during diverging training runs

When comparing two independently trained models, these symmetries can cause their parameter spaces to be misaligned, hindering direct arithmetic operations like skill transfer.

2.2.1 Aligning SwiGLU Layers via Permutation Symmetry

The SwiGLU activation function for a given FFN layer in model i, $\text{Swish}_\beta(x\mathbf{W}_G^{(i)\top}) \odot (x\mathbf{W}_U^{(i)})$, uses gate ($\mathbf{W}_G^{(i)}$) and up-projection ($\mathbf{W}_U^{(i)}$) matrices whose intermediate outputs can be permuted without changing the function, provided the down-projection ($\mathbf{W}_D^{(i)}$) is adjusted accordingly. We find the optimal permutation matrix $\mathbf{P}$ that aligns the weights of two models (1 and 2) by solving the linear assignment problem:

$$\arg\min_{\mathbf{P}\in\mathbb{S}_P} \left(||\mathbf{W}_G^{(1)} - \mathbf{P}\mathbf{W}_G^{(2)}||^2 + ||\mathbf{W}_U^{(1)} - \mathbf{P}\mathbf{W}_U^{(2)}||^2 + ||\mathbf{W}_D^{(1)} - \mathbf{W}_D^{(2)}\mathbf{P}^\top||^2 \right) \tag{3}$$

$$= \arg\max_{\mathbf{P}\in\mathbb{S}_P} \langle \mathbf{P}, \mathbf{W}_G^{(1)}\mathbf{W}_G^{(2)\top} + \mathbf{W}_U^{(1)}\mathbf{W}_U^{(2)\top} + \mathbf{W}_D^{(1)\top}\mathbf{W}_D^{(2)} \rangle. \tag{4}$$

This problem is solved efficiently using the Hungarian algorithm. Model 2 weights are then updated:

$$\mathbf{W}_G^{(2)} \to \mathbf{P}\mathbf{W}_G^{(2)}, \quad \mathbf{W}_U^{(2)} \to \mathbf{P}\mathbf{W}_U^{(2)}, \quad \mathbf{W}_D^{(2)} \to \mathbf{W}_D^{(2)}\mathbf{P}^\top. \tag{5}$$

2.3 Aligning GQA Layers via Rotation and Scaling Symmetry

The absence of element-wise non-linearities between query, key, and value projections in attention layers allows for continuous rotation and scaling symmetries. We align the GQA layers in three steps.

2.3.1 Optimal Rotation for Query and Key Heads

In GQA, all query heads (Q_k) in a group (G_j) share a single key head (K_j). We find a single rotation matrix $\mathbf{R}_{QK_j}$ for this shared space by solving the Orthogonal Procrustes Problem:

$$\underset{\mathbf{R}_{QK_j} \in \mathbb{S}_O}{\arg\min} \left(\sum_{k \in G_j} ||\mathbf{W}_{Q_k}^{(1)} - \mathbf{R}_{QK_j}\mathbf{W}_{Q_k}^{(2)}||^2 + ||\mathbf{W}_{K_j}^{(1)} - \mathbf{R}_{QK_j}\mathbf{W}_{K_j}^{(2)}||^2 \right). \tag{6}$$

The solution is $\mathbf{R}_{QK_j} = U_{QK_j}V_{QK_j}^\top$ from the SVD of $\mathbf{M}_{QK_j} = \sum_{k \in G_j} \mathbf{W}_{Q_k}^{(1)}{\mathbf{W}_{Q_k}^{(2)}}^\top + \mathbf{W}_{K_j}^{(1)}{\mathbf{W}_{K_j}^{(2)}}^\top$. The weights are updated as:

$$\forall k \in G_j : \mathbf{W}_{Q_k}^{(2)} \to \mathbf{R}_{QK_j}\mathbf{W}_{Q_k}^{(2)}, \quad \mathbf{W}_{K_j}^{(2)} \to \mathbf{R}_{QK_j}\mathbf{W}_{K_j}^{(2)}. \tag{7}$$

A similar procedure is used to find the optimal rotation $\mathbf{R}_{VO_j}$ for the shared value (V_j) and corresponding output (O_k) projections.

2.3.2 Optimal Scaling for Query and Key Heads

After rotation, we find a scalar α to align the scales of the query and key matrices, denoted $\mathbf{W}_{Q_k}^{(2)'}$ and $\mathbf{W}_{K_j}^{(2)'}$ post-rotation. The functionality is preserved if one is scaled by α and the other by $1/\alpha$. We find the optimal α by minimizing:

$$\underset{\alpha \in \mathbb{R} \setminus 0}{\arg\min} \left(\sum_{k \in G_j} ||\mathbf{W}_{Q_k}^{(1)} - \alpha\mathbf{W}_{Q_k}^{(2)'}||^2 + ||\mathbf{W}_{K_j}^{(1)} - \frac{1}{\alpha}\mathbf{W}_{K_j}^{(2)'}||^2 \right). \tag{8}$$

We expand this equation, differentiate it with respect to α to find the local extrema and multiply by α^3 to get rid of denominators. This yields a quartic equation in α that can be solved numerically:

$$\alpha^{\mathbf{4}} \sum_{k \in G_j} ||\mathbf{W}_{Q_k}^{(2)'}||^2 - \alpha^{\mathbf{3}} \sum_{k \in G_j} \langle \mathbf{W}_{Q_k}^{(1)}, \mathbf{W}_{Q_k}^{(2)'} \rangle + \alpha \langle \mathbf{W}_{K_j}^{(1)}, \mathbf{W}_{K_j}^{(2)'} \rangle - ||\mathbf{W}_{K_j}^{(2)'}||^2 = 0. \tag{9}$$

The weights are then updated with the optimal α:

$$\forall k \in G_j : \mathbf{W}_{Q_k}^{(2)'} \to \alpha\mathbf{W}_{Q_k}^{(2)'}, \quad \mathbf{W}_{K_j}^{(2)'} \to \frac{1}{\alpha}\mathbf{W}_{K_j}^{(2)'}. \tag{10}$$

3 Results

3.1 Experimental Setup

Our experiment investigates the transfer of specialized reasoning skills across models from parallel development tracks, using 8-billion-parameter models from the Llama 3.1 family. Our setup involves three models originating from a common ancestor, `Llama-3.1-Base`. The first model is the official branch, containing the standard pre-trained `Llama-3.1-Base` and the instruction-tuned `Llama-3.1-Instruct`. The second, a parallel branch, is represented by AI2's Tulu3 series, which was also instruction-tuned from `Llama-3.1-Base` but followed an independent development path. The third group is the skill source: Nvidia's Nemotron model, which was built upon `Llama-3.1-Instruct` to add advanced reasoning capabilities, such as problem deconstruction and self-correction. Our goal is to transfer the specialized reasoning from Nemotron (on the official branch) to Tulu3 (on the parallel branch). To do this, we create a reasoning "skill vector" by subtracting the parameters of `Llama-3.1-Instruct` from those of Nemotron.

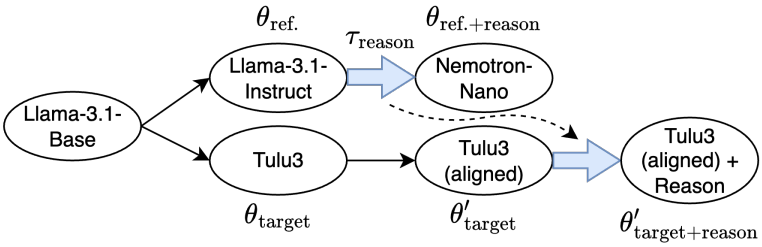

Figure 1: The family tree of models used in our experiments. The diagram illustrates the two divergent fine-tuning branches originating from the common ancestor, `Llama-3.1-Base`. The top branch leads to the reasoning-specialized Nemotron-Nano built on top of `Llama-3.1-Instruct`, while the bottom branch shows the parallel instruction-tuning of the `Tulu3` model. Our work aims to transfer reasoning skills acquired on the top branch to the models on the parallel bottom branch by first aligning the parameter space of `Tulu3` to the one of `Llama-3.1-Instruct`, and then adding the reasoning skill vector.

3.2 Parameter Space Alignment Improves Reasoning Skill Transfer

We evaluate the base models, as well as the `Tulu3` model with the reasoning skills transferred (`+R`) on the following reasoning benchmarks: MATH-500 [12], Minerva-MATH [16], OlympiadBench [10], AMC 2023, AIME 2024 and TheoremQA [4]. We compare adding the reasoning skill vector τ_{reason} to the `Tulu3` model with and without parameter space alignment to the reference `Llama-3.1-Instruct` model. We consider two ways of computing the optimal permutations, rotations and scaling factors, one based on the model activations (`act.align`) and one based on the model weights (`w.align`). We report average results over 8 different random seed evaluations and the standard errors in table 1. We add more details about evaluations in the Appendix.

Table 1: Accuracy and standard error (%) on difficult reasoning benchmarks of the default models and our Tulu3 + Reasoning (with and without neuron matching) models.

Model	MATH-500	Minerva-MATH	OlympiadBench	AMC23	AIME24	TheoremQA	Avg.
`Llama-3.1-Instruct`	45.9±0.39	21.9±0.58	14.1±0.00	28.8±0.72	3.3±0.00	23.9±0.59	30.8
`Tulu3`	39.6±0.45	15.4±0.34	16.1±0.30	24.7±1.92	4.6±0.88	23.6±0.20	29.3
`Nemotron-Nano-v1`	91.4±0.34	53.7±0.44	60.7±0.19	92.5±0.67	58.4±1.99	54.2±0.29	69.7
`Tulu3+R`	83.4±0.51	42.3±1.00	**58.8**±0.28	87.2±1.45	55.9±0.56	49.0±0.45	63.3
`Tulu3(act.align)+R`	**85.8**±0.41	**43.4**±0.79	57.3±0.30	85.6±0.91	56.7±1.79	**50.2**±0.30	63.8
`Tulu3(w.align)+R`	83.8±0.21	42.5±0.78	57.5±0.37	**90.0**±0.67	**61.2**±1.40	49.2±0.30	**64.4**

Firt we observe that the two IF models perform poorly on these complex reasoning tasks, especially when compared to `Nemotron-Nano`. Secondly, with a simple task arithmetic approach we can successfully transfer some of the reasoning skills from `Nemotron-Nano` to the `Tulu3` model, with `Tulu3+R`. performing significantly better than the base `Tulu3`. Lastly, aligning the parameters of `Tulu3` to those of `Llama-3.1-Instruct` improves the reasoning skill transfer even further, with `Tulu3(act.align)+R` and `Tulu3(w.aligned)+R` outperforming `Tulu3+R` on most tasks. Surprisingly, our `Tulu3(w.aligned)+R)` model even outperforms `Nemotron-Nano` on the difficult AIME24 benchmark. We include results on other benchmarks in the Appendix.

4 Conclusion

In this work, we tackle the important challenge of reasoning skill transfer in LLMs, demonstrating that parameter space alignment significantly improves the process. Our main contribution is extending recent alignment methods to support modern architectures with Grouped-Query Attention (GQA) and SwiGLU layers. By aligning models prior to skill transfer via task arithmetic, we show performance gains on difficult reasoning benchmarks. For future work, we plan to extend this analysis to other model families and skills beyond reasoning.

References

[1] J. Ainslie, J. Lee-Thorp, M. de Jong, Y. Zemlyanskiy, F. Lebron, and S. Sanghai. GQA: Training generalized multi-query transformer models from multi-head checkpoints. In *The 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.

[2] S. Ainsworth, J. Hayase, and S. Srinivasa. Git re-basin: Merging models modulo permutation symmetries. In *The Eleventh International Conference on Learning Representations*, 2023.

[3] A. Bercovich, I. Levy, I. Golan, M. Dabbah, R. El-Yaniv, O. Puny, I. Galil, Z. Moshe, T. Ronen, N. Nabwani, I. Shahaf, O. Tropp, E. Karpas, R. Zilberstein, J. Zeng, S. Singhal, A. Bukharin, Y. Zhang, T. Konuk, G. Shen, A. S. Mahabaleshwarkar, B. Kartal, Y. Suhara, O. Delalleau, Z. Chen, Z. Wang, D. Mosallanezhad, A. Renduchintala, H. Qian, D. Rekesh, F. Jia, S. Majumdar, V. Noroozi, W. U. Ahmad, S. Narenthiran, A. Ficek, M. Samadi, J. Huang, S. Jain, I. Gitman, I. Moshkov, W. Du, S. Toshniwal, G. Armstrong, B. Kisacanin, M. Novikov, D. Gitman, E. Bakhturina, P. Varshney, M. Narsimhan, J. P. Scowcroft, J. Kamalu, D. Su, K. Kong, M. Kliegl, R. Karimi, Y. Lin, S. Satheesh, J. Parmar, P. Gundecha, B. Norick, J. Jennings, S. Prabhumoye, S. N. Akter, M. Patwary, A. Khattar, D. Narayanan, R. Waleffe, J. Zhang, B.-Y. Su, G. Huang, T. Kong, P. Chadha, S. Jain, C. Harvey, E. Segal, J. Huang, S. Kashirsky, R. McQueen, I. Putterman, G. Lam, A. Venkatesan, S. Wu, V. Nguyen, M. Kilaru, A. Wang, A. Warno, A. Somasamudramath, S. Bhaskar, M. Dong, N. Assaf, S. Mor, O. U. Argov, S. Junkin, O. Romanenko, P. Larroy, M. Katariya, M. Rovinelli, V. Balas, N. Edelman, A. Bhiwandiwalla, M. Subramaniam, S. Ithape, K. Ramamoorthy, Y. Wu, S. V. Velury, O. Almog, J. Daw, D. Fridman, E. Galinkin, M. Evans, S. Ghosh, K. Luna, L. Derczynski, N. Pope, E. Long, S. Schneider, G. Siman, T. Grzegorzek, P. Ribalta, M. Katariya, C. Alexiuk, J. Conway, T. Saar, A. Guan, K. Pawelec, S. Prayaga, O. Kuchaiev, B. Ginsburg, O. Olabiyi, K. Briski, J. Cohen, B. Catanzaro, J. Alben, Y. Geifman, and E. Chung. Llama-nemotron: Efficient reasoning models, 2025.

[4] W. Chen, M. Yin, M. Ku, P. Lu, Y. Wan, X. Ma, J. Xu, X. Wang, and T. Xia. TheoremQA: A theorem-driven question answering dataset. In H. Bouamor, J. Pino, and K. Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7889–7901, Singapore, Dec. 2023. Association for Computational Linguistics.

[5] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano, C. Hesse, and J. Schulman. Training verifiers to solve math word problems, 2021.

[6] D. Dua, Y. Wang, P. Dasigi, G. Stanovsky, S. Singh, and M. Gardner. DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs. In J. Burstein, C. Doran, and T. Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2368–2378, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[7] L. Gao, J. Tow, B. Abbasi, S. Biderman, S. Black, A. DiPofi, C. Foster, L. Golding, J. Hsu, A. Le Noac'h, H. Li, K. McDonell, N. Muennighoff, C. Ociepa, J. Phang, L. Reynolds, H. Schoelkopf, A. Skowron, L. Sutawika, E. Tang, A. Thite, B. Wang, K. Wang, and A. Zou. The language model evaluation harness, 07 2024.

[8] Z. Gou and Y. Zhang. math-evaluation-harness: A simple toolkit for benchmarking llms on mathematical reasoning tasks. `https://github.com/ZubinGou/math-evaluation-harness`, 2023.

[9] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, A. Yang, A. Fan, A. Goyal, A. Hartshorn, A. Yang, A. Mitra, A. Sravankumar, A. Korenev, A. Hinsvark, A. Rao, A. Zhang, A. Rodriguez, A. Gregerson, A. Spataru, B. Roziere, B. Biron, B. Tang, B. Chern, C. Caucheteux, C. Nayak, C. Bi, C. Marra, C. McConnell, C. Keller, C. Touret, C. Wu, C. Wong, C. C. Ferrer, C. Nikolaidis, D. Allonsius, D. Song, D. Pintz, D. Livshits, D. Wyatt, D. Esiobu, D. Choudhary, D. Mahajan, D. Garcia-Olano, D. Perino, D. Hupkes, E. Lakomkin, E. AlBadawy, E. Lobanova, E. Dinan, E. M. Smith, F. Radenovic, F. Guzmán, F. Zhang, G. Synnaeve, G. Lee, G. L. Anderson, G. Thattai, G. Nail, G. Mialon, G. Pang, G. Cucurell, H. Nguyen, H. Korevaar, H. Xu, H. Touvron, I. Zarov, I. A.

Ibarra, I. Kloumann, I. Misra, I. Evtimov, J. Zhang, J. Copet, J. Lee, J. Geffert, J. Vranes, J. Park, J. Mahadeokar, J. Shah, J. van der Linde, J. Billock, J. Hong, J. Lee, J. Fu, J. Chi, J. Huang, J. Liu, J. Wang, J. Yu, J. Bitton, J. Spisak, J. Park, J. Rocca, J. Johnstun, J. Saxe, J. Jia, K. V. Alwala, K. Prasad, K. Upasani, K. Plawiak, K. Li, K. Heafield, K. Stone, K. El-Arini, K. Iyer, K. Malik, K. Chiu, K. Bhalla, K. Lakhotia, L. Rantala-Yeary, L. van der Maaten, L. Chen, L. Tan, L. Jenkins, L. Martin, L. Madaan, L. Malo, L. Blecher, L. Landzaat, L. de Oliveira, M. Muzzi, M. Pasupuleti, M. Singh, M. Paluri, M. Kardas, M. Tsimpoukelli, M. Oldham, M. Rita, M. Pavlova, M. Kambadur, M. Lewis, M. Si, M. K. Singh, M. Hassan, N. Goyal, N. Torabi, N. Bashlykov, N. Bogoychev, N. Chatterji, N. Zhang, O. Duchenne, O. Çelebi, P. Alrassy, P. Zhang, P. Li, P. Vasic, P. Weng, P. Bhargava, P. Dubal, P. Krishnan, P. S. Koura, P. Xu, Q. He, Q. Dong, R. Srinivasan, R. Ganapathy, R. Calderer, R. S. Cabral, R. Stojnic, R. Raileanu, R. Maheswari, R. Girdhar, R. Patel, R. Sauvestre, R. Polidoro, R. Sumbaly, R. Taylor, R. Silva, R. Hou, R. Wang, S. Hosseini, S. Chennabasappa, S. Singh, S. Bell, S. S. Kim, S. Edunov, S. Nie, S. Narang, S. Raparthy, S. Shen, S. Wan, S. Bhosale, S. Zhang, S. Vandenhende, S. Batra, S. Whitman, S. Sootla, S. Collot, S. Gururangan, S. Borodinsky, T. Herman, T. Fowler, T. Sheasha, T. Georgiou, T. Scialom, T. Speckbacher, T. Mihaylov, T. Xiao, U. Karn, V. Goswami, V. Gupta, V. Ramanathan, V. Kerkez, V. Gonguet, V. Do, V. Vogeti, V. Albiero, V. Petrovic, W. Chu, W. Xiong, W. Fu, W. Meers, X. Martinet, X. Wang, X. Wang, X. E. Tan, X. Xia, X. Xie, X. Jia, X. Wang, Y. Goldschlag, Y. Gaur, Y. Babaei, Y. Wen, Y. Song, Y. Zhang, Y. Li, Y. Mao, Z. D. Coudert, Z. Yan, Z. Chen, Z. Papakipos, A. Singh, A. Srivastava, A. Jain, A. Kelsey, A. Shajnfeld, A. Gangidi, A. Victoria, A. Goldstand, A. Menon, A. Sharma, A. Boesenberg, A. Baevski, A. Feinstein, A. Kallet, A. Sangani, A. Teo, A. Yunus, A. Lupu, A. Alvarado, A. Caples, A. Gu, A. Ho, A. Poulton, A. Ryan, A. Ramchandani, A. Dong, A. Franco, A. Goyal, A. Saraf, A. Chowdhury, A. Gabriel, A. Bharambe, A. Eisenman, A. Yazdan, B. James, B. Maurer, B. Leonhardi, B. Huang, B. Loyd, B. D. Paola, B. Paranjape, B. Liu, B. Wu, B. Ni, B. Hancock, B. Wasti, B. Spence, B. Stojkovic, B. Gamido, B. Montalvo, C. Parker, C. Burton, C. Mejia, C. Liu, C. Wang, C. Kim, C. Zhou, C. Hu, C.-H. Chu, C. Cai, C. Tindal, C. Feichtenhofer, C. Gao, D. Civin, D. Beaty, D. Kreymer, D. Li, D. Adkins, D. Xu, D. Testuggine, D. David, D. Parikh, D. Liskovich, D. Foss, D. Wang, D. Le, D. Holland, E. Dowling, E. Jamil, E. Montgomery, E. Presani, E. Hahn, E. Wood, E.-T. Le, E. Brinkman, E. Arcaute, E. Dunbar, E. Smothers, F. Sun, F. Kreuk, F. Tian, F. Kokkinos, F. Ozgenel, F. Caggioni, F. Kanayet, F. Seide, G. M. Florez, G. Schwarz, G. Badeer, G. Swee, G. Halpern, G. Herman, G. Sizov, Guangyi, Zhang, G. Lakshminarayanan, H. Inan, H. Shojanazeri, H. Zou, H. Wang, H. Zha, H. Habeeb, H. Rudolph, H. Suk, H. Aspegren, H. Goldman, H. Zhan, I. Damlaj, I. Molybog, I. Tufanov, I. Leontiadis, I.-E. Veliche, I. Gat, J. Weissman, J. Geboski, J. Kohli, J. Lam, J. Asher, J.-B. Gaya, J. Marcus, J. Tang, J. Chan, J. Zhen, J. Reizenstein, J. Teboul, J. Zhong, J. Jin, J. Yang, J. Cummings, J. Carvill, J. Shepard, J. McPhie, J. Torres, J. Ginsburg, J. Wang, K. Wu, K. H. U, K. Saxena, K. Khandelwal, K. Zand, K. Matosich, K. Veeraraghavan, K. Michelena, K. Li, K. Jagadeesh, K. Huang, K. Chawla, K. Huang, L. Chen, L. Garg, L. A, L. Silva, L. Bell, L. Zhang, L. Guo, L. Yu, L. Moshkovich, L. Wehrstedt, M. Khabsa, M. Avalani, M. Bhatt, M. Mankus, M. Hasson, M. Lennie, M. Reso, M. Groshev, M. Naumov, M. Lathi, M. Keneally, M. Liu, M. L. Seltzer, M. Valko, M. Restrepo, M. Patel, M. Vyatskov, M. Samvelyan, M. Clark, M. Macey, M. Wang, M. J. Hermoso, M. Metanat, M. Rastegari, M. Bansal, N. Santhanam, N. Parks, N. White, N. Bawa, N. Singhal, N. Egebo, N. Usunier, N. Mehta, N. P. Laptev, N. Dong, N. Cheng, O. Chernoguz, O. Hart, O. Salpekar, O. Kalinli, P. Kent, P. Parekh, P. Saab, P. Balaji, P. Rittner, P. Bontrager, P. Roux, P. Dollar, P. Zvyagina, P. Ratanchandani, P. Yuvraj, Q. Liang, R. Alao, R. Rodriguez, R. Ayub, R. Murthy, R. Nayani, R. Mitra, R. Parthasarathy, R. Li, R. Hogan, R. Battey, R. Wang, R. Howes, R. Rinott, S. Mehta, S. Siby, S. J. Bondu, S. Datta, S. Chugh, S. Hunt, S. Dhillon, S. Sidorov, S. Pan, S. Mahajan, S. Verma, S. Yamamoto, S. Ramaswamy, S. Lindsay, S. Lindsay, S. Feng, S. Lin, S. C. Zha, S. Patil, S. Shankar, S. Zhang, S. Zhang, S. Wang, S. Agarwal, S. Sajuyigbe, S. Chintala, S. Max, S. Chen, S. Kehoe, S. Satterfield, S. Govindaprasad, S. Gupta, S. Deng, S. Cho, S. Virk, S. Subramanian, S. Choudhury, S. Goldman, T. Remez, T. Glaser, T. Best, T. Koehler, T. Robinson, T. Li, T. Zhang, T. Matthews, T. Chou, T. Shaked, V. Vontimitta, V. Ajayi, V. Montanez, V. Mohan, V. S. Kumar, V. Mangla, V. Ionescu, V. Poenaru, V. T. Mihailescu, V. Ivanov, W. Li, W. Wang, W. Jiang, W. Bouaziz, W. Constable, X. Tang, X. Wu, X. Wang, X. Wu, X. Gao, Y. Kleinman, Y. Chen, Y. Hu, Y. Jia, Y. Qi, Y. Li, Y. Zhang, Y. Zhang, Y. Adi, Y. Nam, Yu, Wang, Y. Zhao, Y. Hao, Y. Qian, Y. Li, Y. He, Z. Rait, Z. DeVito, Z. Rosnbrick, Z. Wen, Z. Yang, Z. Zhao, and Z. Ma. The llama 3 herd of models, 2024.

[10] C. He, R. Luo, Y. Bai, S. Hu, Z. Thai, J. Shen, J. Hu, X. Han, Y. Huang, Y. Zhang, J. Liu, L. Qi, Z. Liu, and M. Sun. OlympiadBench: A challenging benchmark for promoting AGI with olympiad-level bilingual multimodal scientific problems. In L.-W. Ku, A. Martins, and V. Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3828–3850, Bangkok, Thailand, Aug. 2024. Association for Computational Linguistics.

[11] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt. Measuring massive multitask language understanding. In *International Conference on Learning Representations*, 2021.

[12] D. Hendrycks, C. Burns, S. Kadavath, A. Arora, S. Basart, E. Tang, D. Song, and J. Steinhardt. Measuring mathematical problem solving with the math dataset. In J. Vanschoren and S. Yeung, editors, *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

[13] D. Hendrycks, C. Burns, S. Kadavath, A. Arora, S. Basart, E. Tang, D. Song, and J. Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

[14] G. Ilharco, M. T. Ribeiro, M. Wortsman, L. Schmidt, H. Hajishirzi, and A. Farhadi. Editing models with task arithmetic. In *The Eleventh International Conference on Learning Representations*, 2023.

[15] N. Lambert, J. Morrison, V. Pyatkin, S. Huang, H. Ivison, F. Brahman, L. J. V. Miranda, A. Liu, N. Dziri, S. Lyu, Y. Gu, S. Malik, V. Graf, J. D. Hwang, J. Yang, R. L. Bras, O. Tafjord, C. Wilhelm, L. Soldaini, N. A. Smith, Y. Wang, P. Dasigi, and H. Hajishirzi. Tulu 3: Pushing frontiers in open language model post-training, 2025.

[16] A. Lewkowycz, A. Andreassen, D. Dohan, E. Dyer, H. Michalewski, V. Ramasesh, A. Slone, C. Anil, I. Schlag, T. Gutman-Solo, Y. Wu, B. Neyshabur, G. Gur-Ari, and V. Misra. Solving quantitative reasoning problems with language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 3843–3857. Curran Associates, Inc., 2022.

[17] D. Rein, B. L. Hou, A. C. Stickland, J. Petty, R. Y. Pang, J. Dirani, J. Michael, and S. R. Bowman. GPQA: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*, 2024.

[18] N. Shazeer. Glu variants improve transformer, 2020.

[19] Z. R. Sprague, X. Ye, K. Bostrom, S. Chaudhuri, and G. Durrett. MuSR: Testing the limits of chain-of-thought with multistep soft reasoning. In *The Twelfth International Conference on Learning Representations*, 2024.

[20] A. Srivastava, A. Rastogi, A. Rao, A. A. M. Shoeb, A. Abid, A. Fisch, A. R. Brown, A. Santoro, A. Gupta, A. Garriga-Alonso, A. Kluska, A. Lewkowycz, A. Agarwal, A. Power, A. Ray, A. Warstadt, A. W. Kocurek, A. Safaya, A. Tazarv, A. Xiang, A. Parrish, A. Nie, A. Hussain, A. Askell, A. Dsouza, A. Slone, A. Rahane, A. S. Iyer, A. J. Andreassen, A. Madotto, A. Santilli, A. Stuhlmüller, A. M. Dai, A. La, A. K. Lampinen, A. Zou, A. Jiang, A. Chen, A. Vuong, A. Gupta, A. Gottardi, A. Norelli, A. Venkatesh, A. Gholamidavoodi, A. Tabassum, A. Menezes, A. Kirubarajan, A. Mullokandov, A. Sabharwal, A. Herrick, A. Efrat, A. Erdem, A. Karakaş, B. R. Roberts, B. S. Loe, B. Zoph, B. Bojanowski, B. Özyurt, B. Hedayatnia, B. Neyshabur, B. Inden, B. Stein, B. Ekmekci, B. Y. Lin, B. Howald, B. Orinion, C. Diao, C. Dour, C. Stinson, C. Argueta, C. Ferri, C. Singh, C. Rathkopf, C. Meng, C. Baral, C. Wu, C. Callison-Burch, C. Waites, C. Voigt, C. D. Manning, C. Potts, C. Ramirez, C. E. Rivera, C. Siro, C. Raffel, C. Ashcraft, C. Garbacea, D. Sileo, D. Garrette, D. Hendrycks, D. Kilman, D. Roth, C. D. Freeman, D. Khashabi, D. Levy, D. M. González, D. Perszyk, D. Hernandez, D. Chen, D. Ippolito, D. Gilboa, D. Dohan, D. Drakard, D. Jurgens, D. Datta, D. Ganguli, D. Emelin, D. Kleyko, D. Yuret, D. Chen, D. Tam, D. Hupkes, D. Misra, D. Buzan, D. C. Mollo, D. Yang, D.-H. Lee, D. Schrader, E. Shutova, E. D. Cubuk, E. Segal, E. Hagerman, E. Barnes, E. Donoway, E. Pavlick, E. Rodolà, E. Lam, E. Chu, E. Tang, E. Erdem, E. Chang, E. A. Chi, E. Dyer,

E. Jerzak, E. Kim, E. E. Manyasi, E. Zheltonozhskii, F. Xia, F. Siar, F. Martínez-Plumed, F. Happé, F. Chollet, F. Rong, G. Mishra, G. I. Winata, G. de Melo, G. Kruszewski, G. Parascandolo, G. Mariani, G. X. Wang, G. Jaimovitch-Lopez, G. Betz, G. Gur-Ari, H. Galijasevic, H. Kim, H. Rashkin, H. Hajishirzi, H. Mehta, H. Bogar, H. F. A. Shevlin, H. Schuetze, H. Yakura, H. Zhang, H. M. Wong, I. Ng, I. Noble, J. Jumelet, J. Geissinger, J. Kernion, J. Hilton, J. Lee, J. F. Fisac, J. B. Simon, J. Koppel, J. Zheng, J. Zou, J. Kocon, J. Thompson, J. Wingfield, J. Kaplan, J. Radom, J. Sohl-Dickstein, J. Phang, J. Wei, J. Yosinski, J. Novikova, J. Bosscher, J. Marsh, J. Kim, J. Taal, J. Engel, J. Alabi, J. Xu, J. Song, J. Tang, J. Waweru, J. Burden, J. Miller, J. U. Balis, J. Batchelder, J. Berant, J. Frohberg, J. Rozen, J. Hernandez-Orallo, J. Boudeman, J. Guerr, J. Jones, J. B. Tenenbaum, J. S. Rule, J. Chua, K. Kanclerz, K. Livescu, K. Krauth, K. Gopalakrishnan, K. Ignatyeva, K. Markert, K. Dhole, K. Gimpel, K. Omondi, K. W. Mathewson, K. Chiafullo, K. Shkaruta, K. Shridhar, K. McDonell, K. Richardson, L. Reynolds, L. Gao, L. Zhang, L. Dugan, L. Qin, L. Contreras-Ochando, L.-P. Morency, L. Moschella, L. Lam, L. Noble, L. Schmidt, L. He, L. Oliveros-Colón, L. Metz, L. K. Senel, M. Bosma, M. Sap, M. T. Hoeve, M. Farooqi, M. Faruqui, M. Mazeika, M. Baturan, M. Marelli, M. Maru, M. J. Ramirez-Quintana, M. Tolkiehn, M. Giulianelli, M. Lewis, M. Potthast, M. L. Leavitt, M. Hagen, M. Schubert, M. O. Baitemirova, M. Arnaud, M. McElrath, M. A. Yee, M. Cohen, M. Gu, M. Ivanitskiy, M. Starritt, M. Strube, M. Swędrowski, M. Bevilacqua, M. Yasunaga, M. Kale, M. Cain, M. Xu, M. Suzgun, M. Walker, M. Tiwari, M. Bansal, M. Aminnaseri, M. Geva, M. Gheini, M. V. T, N. Peng, N. A. Chi, N. Lee, N. G.-A. Krakover, N. Cameron, N. Roberts, N. Doiron, N. Martinez, N. Nangia, N. Deckers, N. Muennighoff, N. S. Keskar, N. S. Iyer, N. Constant, N. Fiedel, N. Wen, O. Zhang, O. Agha, O. Elbaghdadi, O. Levy, O. Evans, P. A. M. Casares, P. Doshi, P. Fung, P. P. Liang, P. Vicol, P. Alipoormolabashi, P. Liao, P. Liang, P. W. Chang, P. Eckersley, P. M. Htut, P. Hwang, P. Miłkowski, P. Patil, P. Pezeshkpour, P. Oli, Q. Mei, Q. Lyu, Q. Chen, R. Banjade, R. E. Rudolph, R. Gabriel, R. Habacker, R. Risco, R. Millière, R. Garg, R. Barnes, R. A. Saurous, R. Arakawa, R. Raymaekers, R. Frank, R. Sikand, R. Novak, R. Sitelew, R. L. Bras, R. Liu, R. Jacobs, R. Zhang, R. Salakhutdinov, R. A. Chi, S. R. Lee, R. Stovall, R. Teehan, R. Yang, S. Singh, S. M. Mohammad, S. Anand, S. Dillavou, S. Shleifer, S. Wiseman, S. Gruetter, S. R. Bowman, S. S. Schoenholz, S. Han, S. Kwatra, S. A. Rous, S. Ghazarian, S. Ghosh, S. Casey, S. Bischoff, S. Gehrmann, S. Schuster, S. Sadeghi, S. Hamdan, S. Zhou, S. Srivastava, S. Shi, S. Singh, S. Asaadi, S. S. Gu, S. Pachchigar, S. Toshniwal, S. Upadhyay, S. S. Debnath, S. Shakeri, S. Thormeyer, S. Melzi, S. Reddy, S. P. Makini, S.-H. Lee, S. Torene, S. Hatwar, S. Dehaene, S. Divic, S. Ermon, S. Biderman, S. Lin, S. Prasad, S. Piantadosi, S. Shieber, S. Misherghi, S. Kiritchenko, S. Mishra, T. Linzen, T. Schuster, T. Li, T. Yu, T. Ali, T. Hashimoto, T.-L. Wu, T. Desbordes, T. Rothschild, T. Phan, T. Wang, T. Nkinyili, T. Schick, T. Kornev, T. Tunduny, T. Gerstenberg, T. Chang, T. Neeraj, T. Khot, T. Shultz, U. Shaham, V. Misra, V. Demberg, V. Nyamai, V. Raunak, V. V. Ramasesh, vinay uday prabhu, V. Padmakumar, V. Srikumar, W. Fedus, W. Saunders, W. Zhang, W. Vossen, X. Ren, X. Tong, X. Zhao, X. Wu, X. Shen, Y. Yaghoobzadeh, Y. Lakretz, Y. Song, Y. Bahri, Y. Choi, Y. Yang, S. Hao, Y. Chen, Y. Belinkov, Y. Hou, Y. Hou, Y. Bai, Z. Seid, Z. Zhao, Z. Wang, Z. J. Wang, Z. Wang, and Z. Wu. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*, 2023. Featured Certification.

[21] G. Team, T. Mesnard, C. Hardin, R. Dadashi, S. Bhupatiraju, S. Pathak, L. Sifre, M. Rivière, M. S. Kale, J. Love, P. Tafti, L. Hussenot, P. G. Sessa, A. Chowdhery, A. Roberts, A. Barua, A. Botev, A. Castro-Ros, A. Slone, A. Héliou, A. Tacchetti, A. Bulanova, A. Paterson, B. Tsai, B. Shahriari, C. L. Lan, C. A. Choquette-Choo, C. Crepy, D. Cer, D. Ippolito, D. Reid, E. Buchatskaya, E. Ni, E. Noland, G. Yan, G. Tucker, G.-C. Muraru, G. Rozhdestvenskiy, H. Michalewski, I. Tenney, I. Grishchenko, J. Austin, J. Keeling, J. Labanowski, J.-B. Lespiau, J. Stanway, J. Brennan, J. Chen, J. Ferret, J. Chiu, J. Mao-Jones, K. Lee, K. Yu, K. Millican, L. L. Sjoesund, L. Lee, L. Dixon, M. Reid, M. Mikuła, M. Wirth, M. Sharman, N. Chinaev, N. Thain, O. Bachem, O. Chang, O. Wahltinez, P. Bailey, P. Michel, P. Yotov, R. Chaabouni, R. Comanescu, R. Jana, R. Anil, R. McIlroy, R. Liu, R. Mullins, S. L. Smith, S. Borgeaud, S. Girgin, S. Douglas, S. Pandya, S. Shakeri, S. De, T. Klimenko, T. Hennigan, V. Feinberg, W. Stokowiec, Y. hui Chen, Z. Ahmed, Z. Gong, T. Warkentin, L. Peran, M. Giang, C. Farabet, O. Vinyals, J. Dean, K. Kavukcuoglu, D. Hassabis, Z. Ghahramani, D. Eck, J. Barral, F. Pereira, E. Collins, A. Joulin, N. Fiedel, E. Senter, A. Andreev, and K. Kenealy. Gemma: Open models based on gemini research and technology, 2024.

[22] Y. Wang, X. Ma, G. Zhang, Y. Ni, A. Chandra, S. Guo, W. Ren, A. Arulraj, X. He, Z. Jiang, T. Li, M. Ku, K. Wang, A. Zhuang, R. Fan, X. Yue, and W. Chen. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 95266–95290. Curran Associates, Inc., 2024.

[23] P. Yadav, D. Tam, L. Choshen, C. Raffel, and M. Bansal. TIES-merging: Resolving interference when merging models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[24] L. Yu, B. Yu, H. Yu, F. Huang, and Y. Li. Language models are super mario: Absorbing abilities from homologous models as a free lunch. In R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 57755–57775. PMLR, 21–27 Jul 2024.

[25] B. Zhang, Z. Zheng, Z. Chen, and J. Li. Beyond the permutation symmetry of transformers: The role of rotation for model fusion. In *Forty-second International Conference on Machine Learning*, 2025.

[26] J. Zhou, T. Lu, S. Mishra, S. Brahma, S. Basu, Y. Luan, D. Zhou, and L. Hou. Instruction-following evaluation for large language models, 2023.

A Experimental details

All evaluations of the models were ran on one p4d.24xlarge node with 8xA100 GPUs with 40GB of VRAM each. We used versions of EleutherAI's LM Evaluation Harness [7] for the evaluation on non-reasoning tasks and the Math Evaluation Harness [8] for reasoning tasks. We evaluate the models on the reasoning tasks with temperature of 1.

Data used for activation-based alignment To align the `Tulu3` model parameter space to that of `Llama-3.1-Instruct` using activations, we feed the model the prompts from the following datasets: MMLU (validation set) [11], DROP (validation set) [6], MATH (dev. set) [13], GSM8K (training set) [5], Tulu-3 SFT Personas Code (train set) and Tulu-3 WildChat IF On Policy (train set) [15]. We only extract the activations of the input prompts, stopping the activation extraction before generation starts.

B Ablation study

We also perform an ablation to see which symmetries make the biggest impact when accounted for during the matching process, the results are shown in table 2. First we note that in our experiments, the optimal permutations found through the weight alignment process are always the identity (i.e. no neuron permutations). We hypothesize this is because the models haven't diverged enough during their separate training procedures to warrant the need for a hard re-ordering of any neurons. We see that the biggest gain in accuracy comes from accounting for rotations. Scaling the Q and K layers also brings some minor improvements beyond the non aligned reasoning transfer model.

Table 2: Accuracy and standard error (%) on difficult reasoning benchmarks for different symmetries taken into consideration

Model	MATH-500	Minerva-MATH	OlympiadBench	AMC23	AIME24	TheoremQA	Avg.
`Tulu3+R`	83.4±0.51	42.3±1.00	58.8±0.28	87.2±1.45	55.9±0.56	49.0±0.45	63.3
`Tulu3(act.align rot.)+R`	84.1±0.45	42.7±1.22	58.1±0.31	91.9±1.22	58.7±2.43	48.7±0.24	64.5
`Tulu3(act.align scale)+R`	84.0±0.64	40.9±0.56	58.7±0.39	87.5±1.83	57.9±1.40	49.3±0.45	63.6
`Tulu3(w.align)+R`	83.8±0.21	42.5±0.78	57.5±0.37	90.0±0.67	61.2±1.40	49.2±0.30	64.4

C Evaluations on other benchmarks

We also evaluate our models on other benchmarks: IFEval [26], MMLU Pro [22], GPQA [17], MUSR [19] and BBH [20]. The results are shown in table 3. We use greedy sampling for these benchmarks, therefore we report only a single run of evaluations. We note a drop in accuracy on these tasks when transferring reasoning skills to the `Tulu3` model. This is somewhat expected since task arithmetic approaches are known to suffer from negative interference leading to worst performance on some tasks than the performance of the original models [14, 23, 24]. We note that aligning the `Tulu3` parameter space to that of `Llama-3.1-Instruct` doesn't cause any further drops in performance on these benchmarks, instead we observe minimal gains. Interestingly, it seems that activation alignment (`Tulu3(act.align)+R`) specifically behaves in a different fashion than both no alignment (`Tulu3+R`) and weight alignment (`Tulu3(w.align)+R`). `Tulu3(act.align)+R` preserves the strong instruction following skills from the base models while performing worse on some tasks such as MMLU Pro and BBH.

Table 3: Accuracy (%) on non-reasoning benchmarks.

Model	IFEval	MMLU Pro	GPQA	MUSR	BBH	Avg.
`Llama-3.1-Instruct`	79.3	36.4	32.7	37.8	48.5	47.0
`Tulu3`	83.5	28.3	29.5	43.0	48.0	46.4
`Nemotron-Nano-v1`	79.0	31.9	31.1	33.9	46.0	44.4
`Tulu3+R`	59.7	29.2	27.4	37.3	41.3	39.0
`Tulu3(act.align)+R`	80.1	12.5	28.3	40.7	35.0	39.3
`Tulu3(w.align)+R`	60.1	29.2	27.9	37.3	41.3	39.2